System Demonstration

Topic Maps-based Semblogging
with semblog-tm

Alexander Sigel, Lars Bünten & Marcus Frank
University of Cologne, Germany
Dept. of Information Systems & Information Management
sigel@asigel.de | lars.b@gmx.de
Semantic blogging (semblogging) crosses blogs with the Semantic Web to enrich blog entries with more explicit and machine-understandable metadata, relating both to structure and content.

This semantics is conveyed by semantic tags (semtags), connected by semantic associations (semrels).
0. About

Semblog-tm is a web-based application, realized as a plugin for the blojsom blogging engine, using Java and Tomcat with Velocity templates, with topic map operations via TMAPI, using TM4J and OKS as topic map engines, and with Axis for web services. A demonstrator is publicly available, and the source can be retrieved from sourceforge (semblog.sf.net).
Outline

1. Motivation, Problem and Diagnosis
2. Prior art and related work
3. Approach
4. Demonstration of main functionalities as seen by sembloggers
5. System architecture
6. Discussion and Implications
Simple tagging is not enough
Improve conventional blog tagging

→ Folksonomies
Lacks explicit structure (structural metadata)
Lacks explicit semantics (content metadata)
  ▶ Tags not semantically grounded (only names/strings)
  ▶ Tags not interrelated by typed semantic relations
Difficult to syndicate different blog entries by subject
Conventional blog tagging → Folksonomies
► Tags not semantically grounded
► Tags not interrelated by typed semantic relations

Enrich conventional blogging with more explicit, semantic, and machine-understandable metadata, relating both to structure and content
1. Motivation, Problem, and Diagnosis

- Structured tagging exists
- Semantic wikis exist
- Semantic blogging with RDF existed, but not with Topic Maps
  - we were curious and wished to replicate
  - we believe that modelling the association types with topic maps is more natural, and the semantic retrieval is more powerful)

Challenge:
How to make semblogging easy and effortless for users?
2. Prior art and related work (RDF)

1. **Seminal semblog system** (by Steve Cayzer and colleagues, HP/SWAD-E)
2. **Haystack semblogging client** (Karger & Quan)
3. **Semblog** (Personal Knowledge Publishing Suite) (Ohmukai, Takeda, Numa and colleagues)
4. **Semblogging research project** (KMI: Sereno, Eisenstadt, Buckingham Shum)
5. **Tagsocratic project** (Avesani, Cova, Hayes, Massa)
6. **semiBlog** (semblogging to publish desktop data)
2. Prior art and related work (Topic Maps)

Early ideas:
- Jack Park
- Dmitry Bogachev
- A playground implementation with OKS kindly provided by Lars Marius Garshol

Recently:
- Rajbhandari, Andres, Naito & Wuwongse on semantic-augmented support in spatial-temporal multimedia blog management
Our analysis of the **six semblogging approaches and systems** identified reveals that:

► all are RDF-based
► suffer from not using Published Subjects as proxies for subjects, and
► do not employ semantic relations motivated by knowledge organization.

When we started in 2005, there was **no open effort on a topic maps-based semblogging system, and no research whatsoever** (see open space TMRA05 report)
3. Approach

We introduce semblog-tm, a

- Topic Maps- and
- PSI-based semblogging system,

- whose set of semantic relations follows a proposal in knowledge organization for the organization of association types
Starting from an application scenario with several sembloggers and servers, basic requirements are derived from the **three main use cases**:

1. **Managing lightweight ontologies**: Defining ontologies (semtags connected by semantic associations).
2. **Semblogging**: Connecting semblog entries with entries from those ontologies to attach semantics.
3. **Managing aggregated semblogging data**: Provide other semantically-enabled systems with topic map fragments.
4. Demonstration

We demonstrate and discuss typical capabilities of our prototype implementation, as seen from the users' perspective.
Donnerstag, 28 September 2006
Attending TMRA06 (1)

After some problems with my flight, I finally arrived in Leipzig to attend TMRA06, and made my way to Villa Ida. Nice new building.
(…)

In the spirit of this year's motto, Leveraging the Semantics, Steve Pepper proposed Public Resource Identifiers (PRIs), the successor of PIs, to aid semantic interoperability.
(…)

In the demonstration session, a topic maps-based semblogging system called semblog was shown.
(…)

Posted by semblogger, Nick at 2:50 PM in 4
Comments (0) Trackbacks (0)

This blog entry has the following semantics (xml):
* is_written_by semblogger, Nick

Formal metadata
Nick discovering existing PSIs

Discover a PSI on a webpage (here: provided via Topincs)
#3 Nick discovering other blog entries
#4 Nick discovering semantics in other semblogger's blogs
#5 Nick discovering semtags in Lutz Maicher's blog

Follow from blog to semtag!

Semantic metadata

Leveraging Semantics, Topic Maps, and Information Retrieval

Leveraging Tags and Information Retrieval

Leveraging Semantic Concepts in Other Blogs

Leveraging Semantic Concepts in Other Blogs

Leveraging Semantic Concepts in Other Blogs

Leveraging Semantic Concepts in Other Blogs
#6 Nick following semtags in Lutz Maicher's blog

Blog of Lutz Maicher
This is my personal blog

Semtag (concept)
Semtag-semtag relations

Ontology level:

Topic Maps-based Semblogging
#6 Nick following semtags in Lutz Maicher's blog

Follow to related semtag!
#7 Nick following semtags in student's blog
#7 Nick following semtags in student's blog

Three semtags:

Villa Ida
Leipzig
Zentrum für ...

Students of the University of Leipzig
Syndicated blog by students of the University of Leipzig
Back to main page
Associate this topic with another topic

Topic Names:
- Villa Ida
  - Scope: semblogger:location
  - sort name: Villa Ida

Topic Types:
- topic

Topic Instances:
- Association:
  - a subject of Villa Ida
  - is a place for its geographic part of Leipzig

Occurrences:
200609211118 (Type: publication date)
"Villa Ida" is a historical building in Leipzig. The Villa Ida is surrounded by the Mediencampus Leipzig (Campus Villa Ida), an institutional part of the Leipzig Media Foundation. (Type: human abstractive)
http://www.leipzig-mediencampus.de/Ida/ (Type: primary topic subject identifier)
(Type: machine metadata)

Subject Identifiers:

Leipzig

Topic Names:
- Leipzig
  - Scope: semblogger:location
  - sort name: Leipzig

Topic Types:
- topic

Topic Instances:
- Association:
  - is in place of its geographic part of Germany
  - is a place for its geographic whole part University Leipzig
  - is a place for its geographic whole part Villa Ida
  - is a place for its geographic whole part Zentrum für Information-, Wissens- und Dienstleistungsmangement

Occurrences:
20060921106 (Type: publication date)
http://blaleiste.9080/sem/Leapizg (Type: primary topic subject identifier)
The town Leipzig in Germany. (Type: human abstractive)
(Type: machine metadata)

Subject Identifiers:

Zentrum für ...

Topic Names:
- Zentrum für Information-, Wissens- und Dienstleistungsmangement
  - Scope: semblogger:location
  - sort name: Zentrum für Information-, Wissens- und Dienstleistungsmangement

Topic Types:
- topic

Topic Instances:
- Association:
  - is a cooperation with University Leipzig
  - is subject of What is the Zentrum für Information-, Wissens- und Dienstleistungsmangement
  - is in place of its geographic part of Leipzig

Occurrences:
20060921107 (Type: publication date)
http://blaleiste.9080/sem/Leapizg (Type: primary topic subject identifier)
Zentrum für Information-, Wissens- und Dienstleistungsmangement (Type: human abstractive)
(Type: machine metadata)

Subject Identifiers:
Donnerstag, 28 September 2006

On PRIs and PSIs

A PRI (Public Resource Identifier) is an http: IRI that has been minted for the express purpose of identifying some subject.

PRIs are the successors of PSIs (Published Subjects).

Published Subjects are the successors of Public Subjects.

At TMRA06 I'll present "A Proposal for Public Resource Identifiers."

Here is the abstract:

The only way to achieve semantic interoperability is by making it possible to express what subjects a piece of information is about in such a way that it can be connected to other pieces of information that are about the same subject. Because of the fragility of names, this requires the use of globally unique identifiers for subjects of discourse. This paper describes the requirements on a mechanism for defining and assigning unique global identifiers for arbitrary subjects on the World Wide Web and presents the case for Public Resource Identifiers (PRIs) as the best solution to this problem.

Posted by semblogs, Pepper at 1:20 PM in 
Comments (0)  Trackbacks (0)

This blog entry has the following semantics (edi):

- is about Published Subject
- is about Published Subject Indicator
- is about PRI
- is about TMRA-06
- is about Published Subject Identifier
- is about PSI
- is about Public Subject

2006-10-11  Topic Maps-based Semblogging
#8 Nick following semtags in Steve Pepper's blog (1)

## Three semtags:

<table>
<thead>
<tr>
<th>Public Resource Identifier</th>
<th>Published Subject</th>
<th>Public Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic Names:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Public Resource Identifier</td>
<td>Steve Pepper</td>
<td></td>
</tr>
<tr>
<td><strong>Type:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type Instances:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Issue</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Associations:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- is a subresource of Public Subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occurrences:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://localhost:8080/axis/ps/Public_Resource_Identifier">http://localhost:8080/axis/ps/Public_Resource_Identifier</a> (Type: primary topic subject identifier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200609281325 (Type: publication date)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Type: machine metadata)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topic Names:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Published Subject</td>
<td>Steve Pepper</td>
<td></td>
</tr>
<tr>
<td><strong>Type:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type Instances:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Issue</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Associations:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- is a subresource of Public Subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occurrences:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://localhost:8080/axis/ps/Published_Subject">http://localhost:8080/axis/ps/Published_Subject</a> (Type: primary topic subject identifier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200609281322 (Type: publication date)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Type: machine metadata)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topic Names:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Public Subject</td>
<td>Steve Pepper</td>
<td></td>
</tr>
<tr>
<td><strong>Type:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type Instances:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Issue</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Associations:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- is a subresource of Public Subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occurrences:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://localhost:8080/axis/ps/Public_Subject">http://localhost:8080/axis/ps/Public_Subject</a> (Type: primary topic subject identifier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200609281321 (Type: publication date)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Type: machine metadata)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subject Identifiers:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://localhost:8080/axis/ps/Public_Subject">http://localhost:8080/axis/ps/Public_Subject</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
#10 Nick editing semtags

Blog of Nick Name
My personal blog

Donnerstag, 28 September 2006
Attending TMRA06 (1)

After some problems with my flight, I finally arrived in Leipzig to attend TMRA06, and made my way to Villa ids. Nice new building.

(...) In the spirit of this year's motto, Leveraging the Semantics, Steve Peppers proposed Public Resource Identifiers (PRIs), the successor of PSL, to aid semantic interoperability.

(...) In the demonstration session, a topic-maps-based semblogging system called semblog was shown.

(...) Posted by semblogger Nick at 4:16 PM in

This blog entry has the following semantic tags:
* is written by semblogger Nick

Choose Semtags

User: semblogger_Nick
Blog entry: Attending TMRA06 (1)

Available Semtags (add new): [Check a semtag to add it to this entry]

Selected Semtags: [Check a semtag to delete it from this entry]

All Semtags: [Check a semtag to copy it to the topic-map]

Semtags from last query result: [Check a semtag to copy it to the topic-map]

Edit semantics!
#11 Nick copying and attaching semtags

Select existing semtags
Copy to own view

Copied semtags are now available in own view
#11 Nick copying and attaching semtags

Select from available semtags in own view

Selected semtags now attached to semblog entry
After some problems with my flight, I finally arrived in Leipzig to attend TMRa06, and made my way to Villa Ida. Nice new building.

...)

In the spirit of this year’s motto, Leveraging the Semantics, Steve Pepper proposed Public Resource Identifiers (PRIs), the successor of PSIs, to aid semantic interoperability.

...)

In the demonstration session, a topic maps-based semblogging system called semblog-tm was shown.

...)

Posted by semblogger Nick at 4:35 PM in /
Comments (0) Trackbacks (0)

This blog entry has the following semantics:
- is about TMRa06
- is written by semblogger Nick
- is about PSI
- is about Public Resource Identifier
- is about Leipzig
- is about Villa Ida
- is about Leveraging the Semantics

Display semtags attached to this semblog entry
#13 Nick defining his own semtag (1)
#14 Nick defining his own semtag (2)

ADD NEW SEMTAG

Topic Name: semblog
Topic Sort Name: semblog
Subject Identifier: a system for blogging with semantics

psi automatically created
#15 Nick adding his own semtag to his blog entry

Attach semtag to semblog entry
#18 Nick associating semtags

Associate this semtag with another semtag!

Select association type from predefined list

Select 2nd semtag
#18 Nick associating semtags

The chosen mapping is:

Semblog-tm is beneficial for Leveraging the Semantics
#21 Narrator explaining semantic retrieval
#22 Nick querying semtags by all sembloggers on the same semblog server

Blog of Nick Name
My personal blog

Back to main page

Available semtags for all users [last, semblogger, nick, default, semblogger, username, user]

- Germany
- latest
- Last tagging the semantics
- SI
- Public Resource Identifier
- semtag
- semblog
- TMRA
- TMRA-06
- University Leipzig
- UDK
- Zentral für Information - Wissens- und Umweltmanagementsystem

Query

Follow result
#24 Nick querying semtags by **all sembloggers on all known semblog servers** for the semtag TMRA06 (display complete results)

Available semtags for all users [topicmap.3, topicmap.2, lars, semblogger_nick, default, semblogger_kiel2006, user1]

- Conference
- Conference session
- Demonstrations (Conference session)
- Germany
- Leh3500
- Leveraging the semantics
- PSI
- Public Resource Identifier
- Public Subject
- Published Subject
- Published Subject Identifier
- Published Subject Indicator
- Semantic interoperability
- semblog
- semblogger
- semblogger-tm
- TMRA
- TMRA 06
- TMRA06
- PSI

**Query results for TMRA 06:**

1. Attending TMRA06 (1)
2. On PRIs and PSI
3. TMRA06 Conference

Follow result
#25 Nick displaying the results and exporting them to .xtm

The result

The XTM export

A PRI (Public Resource Identifier) is an https://iri that has been invented for the express purpose of identifying some subject. PRIls are the successors of PSls (Published Subjects). Published Subjects are the successors of Public Subjects.

At TMRA06 I’ll present “A Proposal for Public Resource Identifiers” Here is the abstract.
The only way to achieve semantic interoperability is by making it possible to express what subjects a piece of information is about in such a way that it can be connected with other pieces of information that are about the same subject. Because of the fragility of names, this requires the use of globally unique identifiers for subjects of discourse. This paper describes the requirements on a mechanism for defining and assigning unique global

The XTM export

...
The main components of our system architecture are:

- blogging,
- topic map,
- and services:
  - PSI,
  - knowledge, and
  - registry services (implemented as web services).
#26 Narrator explaining the **general system architecture**
#27 Narrator explaining **scalability and distribution/P2P-like aspects**
#28 Narrator explaining **semantic knowledge services**

- **Registry-Service**
  - GetURIS
  - RegistrarURI
  - remove_non_working_URI

- **PSI-Service**
  - psi_exists
  - create_psi(name)
  - create_psi(name, URI)
  - create_psi_blojson

- **Knowledge-Service**
  - MergeTopicmapFragments
  - get_topics-associated_with_topic
  - get_new_topics-associated_with_topic
  - get_topics_instance_of_topic
  - get_new_topics_instance_of_topic
  - get_topics_with_name

- **Topic Map-Fragment**

---

2006-10-11  Topic Maps-based Semblogging  39
6. Discussion and Implications: Weaknesses

- Currently only a university **research prototype**, not in practical use, needs to move to being actively used by a community.
- Needs some **clean-up** (e.g. GUI, code, naming restrictions, use of published subjects), and optimization.
- Just one particular set of **predefined association types** (clearly needs refinement, e.g. rhetorical ones).
- No possibility to define **application-specific subtypes of association types**.
- By design, **only semtags can be copied**, not semrels (associations).
6. Discussion and Implications: Weaknesses

- Instead of the association type **Synonymy** basename/scope would be more topic mappish
- **PSI service** could be more sophisticated, in the long run to be replaced by a third-party solution
- Use of **TMAPI limits** (e.g. no TMRAP, because TMRAP or tolog via TMAPI is not yet implemented)
- Needs **OKS** for tolog, because TM4J‘s tolog is not sufficient
6. Discussion and Implications: Strengths (USP)

- First **open-source topic maps-based semblogging** system
- The only semblogging system **using PSIs**
- The only semblogging system based on **association types** motivated and governed by Knowledge Organization theory
- First semblogging system with **open semantic knowledge services**, aggregating the data and providing them back to any topic map fragment consumers → cooperation with other PSI services!
6. Discussion and Implications:
Implications for Research and Practice, Outlook

- Expand actual use in semblogging communities
- Explore interoperability with other PSI and topic map aggregation services, with knowledge service consumers
- Explore P2P-like distribution further: e.g. TMRAP and semantic overlay in P2P networks with topic map data
- Further work under way on RDF/Sparql with topic maps/tolog interoperability in Semantic Knowledge Services, consuming semblog-tm services
More Information

► Have a look at the detailed video
System demonstration (@ TMRA06): Video (.avi) (16'40'') and edited transcript (.pdf)
http://www.wim.uni-koeln.de/uploads/media/semblog-tm-demo-at-TMRA06.zip

► Get the source via SVN checkout from semblog.sf.net

► Ask for a free CD

► Read the accompanying paper
(TMRA06 proceedings, forthcoming in 2007)